Supplementary Information

Clonal populations of a human TNBC model display significant functional heterogeneity and divergent growth dynamics in distinct environments

Hendrik J. Kuiken, Sabin Dhakal, Laura M. Selfors, Chandler M. Friend, Tian Zhang, Maurizio Callari, Ron C.J. Schackmann, G. Kenneth Gray, Jett Crowdis, Hyo-eun C. Bhang, Timour Baslan, Frank Stegmeier, Steven P. Gygi, Carlos Caldas, and Joan S. Brugge

Table of contents

Supplementary Discussion

Transcriptional heterogeneity among SCPs of MDA-MB-468
The systematic characterization of SCPs derived from the parental MDA-MB-468 cell line revealed both genetic and transcriptional heterogeneity. Though variation in gene expression was observed among the SCPs, their gene expression profiles correlate highly and more to each other and that of the parental cell line than distinct TNBC cell lines correlate to each other based on their published transcriptional profiles (Supplementary Fig. 1F) (1). However, the correlation coefficient for the gene expression profile of the parental cell line and the previously reported gene expression profile of MDA-MB-468 was lower than observed for SCP pairs and comparable with the correlation coefficients obtained from pairwise comparisons between the published gene expression profiles of distinct TNBC cell lines. This discordance between the MDA-MB-468 gene expression profiles may be the result of different culture conditions, i.e. the parental cell line and SCPs were propagated in culture medium based on DMEM, while the TNBC cell lines that were used in the study of Klijn and colleagues were propagated in culture medium based on RPMI (1).

Single cells of MDA-MB-468 form distinct clusters based on their gene expression profiles
The scRNA-seq analysis of individual cells isolated from the parental cell line provides evidence for distinct cell states or tumor cell subpopulations in the MDA-MB-468 cell line (Fig. 1D). Some of the distinct single cell clusters were enriched for cells with gene expression profiles that correlate with one or more individual SCPs, suggesting that these SCPs and clusters are related, possibly as a result of shared genetic alterations, epigenetic programming, or enrichment for cells with similar transcriptional states. However, the expression profiles of these single cells and SCPs correlate to a lesser extent than the gene expression profiles of SCPs correlate to each other (Supplementary Fig. 1F). This notable difference in correlation coefficients is most likely the result of the subset of genes that were used in these two correlation analyses (925 genes in the scRNA-seq analysis and 12,052 genes in the bulk RNAseq analysis) but could also be explained by the differences between the two sequencing technologies, including sample preparation.

DNA barcode analyses of SCP mixtures propagated *in vitro* indicate a possible role for clonal interactions
The SCPs displayed minimal variation in morphology and growth rate (Fig. 2A-B). The latter can be explained by long-term propagation of the parental MDA-MB-468 cell line prior to establishment of the SCPs, which would select for clones with the highest cellular fitness and most optimal growth rate. However, the DNA barcode analyses employed in this study showed that SCP mixtures propagated as a monolayer using standard culture conditions had notably different growth dynamics than predicted *in silico* based on SCP PDT (Fig. 3C and Supplementary Fig. 3C-D). It is possible that the discordance between the predicted and observed growth dynamics are due to cooperative and competitive interactions among SCPs. In support of this hypothesis is the observation that the growth dynamics of the SCP mixtures at the later time-points bear resemblance to a state of equilibrium with no further enrichment of

3

SCP13 (Fig. 3C and Supplementary Fig. 3C). Alternatively, this difference in growth dynamics may be due to different sensitivities of the clonal populations to repeated passaging of SCP mixtures, involving numerous 5 to 10-minute treatments with trypsin, short-term incubation in suspension, and seeding of cells at reduced densities. Thus, future studies of SCP dynamics are required and may provide insight into the mechanisms by which distinct tumor cell subpopulations cooperate. For example, clonal interactions have been implicated in tumor development and metastasis (2–8).

Divergent evolution of heterogeneous tumor cell populations in distinct contexts
The DNA barcode analyses also revealed that SCP mixtures undergo diverging growth dynamics when propagated using different culture conditions. For example, the composition of SCP mixtures propagated in culture media based on RPMI was distinct from those propagated in culture media based on DMEM (Supplementary Fig. 3E). The divergence of the SCP mixtures propagated in DMEM or RPMI media can potentially be explained by the difference in the concentrations of media components. For example, DMEM has higher concentrations of glucose (4500 mg/L versus 2000 mg/L in RPMI) and essential amino acids compared to RPMI, and several studies have reported that this affects proliferation of both normal and cancer cells (9,10). In addition, RPMI contains glutathione (GSH), which is absent in DMEM. GSH is the most abundant antioxidant in cells and while it is not clearly understood whether GSH can be transported into the cells, intracellular GSH has been shown to have both pro- as well as anti-tumorigenic roles in cancer biology (11,12). The change in base media also caused a reduction in the growth rate of the SCP mixtures (Supplementary Fig. 3F), suggesting that the SCPs underwent adaptation to the new media composition. On the other hand, the mere difference in SCP distribution may have contributed to the reduction in growth rate of the mixtures. In contrast to these monolayer cultures, suspension cultures of SCP mixtures were not enriched for SCP13 (Fig. 3F). Instead, these SCP mixtures displayed temporal enrichment for SCP18, SCP26 and SCP29, none of which were able to form colonies in soft agar particularly well compared to the other SCPs (Fig. 2C). Thus, this observation suggests that clonal interactions may support their survival and proliferation in suspension cultures.

Heterogeneity in SCP tumor growth rate does not correlate with the proportion of TICs
In comparison to the SCP mixtures propagated *in vitro*, xenograft tumors of SCP mixtures displayed a unique enrichment profile with a strong increase in SCP32 abundance over time. When injected individually, SCPs formed tumors of variable size and mass and in particular SCP32 formed tumors that were much larger in size and mass than those derived from other SCPs or the parental cell line. The heterogeneity in SCP tumor mass is not explained by differences in the proportion of TICs, characterized by high levels of CD44 and low levels of CD24 (Supplementary Fig. 5A-C) (13). In addition, the CyTOF analysis showed that SCP15 and SCP28 have a significantly higher fraction of CD44^highCD24^low cells than the parental cell line (p = 0.0019 and p = 0.0142, respectively). However, SCP15 did not form tumors in NOD/Scid mice and SCP28 tumors were not significantly larger than tumors derived from the parental cell line (Fig 4A and Supplementary Fig. 4A-C). It is possible that additional markers are required for the identification of TICs in this model. For example, there is evidence that epithelial specific antigens may aid in the identification of TICs in breast cancer cell lines (14). Moreover, the lack

4

of correlation between SCP tumor mass and the proportion of CD44^{high}CD24^{low} cells in the SCP cultures does not exclude the possibility that the observed differences the proportion of CD44^{high}CD24^{low} cells correlates with the ability of the SCPs to form xenograft tumors at low cell numbers as is common practice in experiments designed to test for enrichment of TICs.

Type I interferon signaling functions as a tumor suppressor in SCP tumors
Further characterization of SCP tumors revealed that SCP32 tumors were characterized by reduced expression of IFN target genes (Fig. 4D-H). In addition, analysis of SCP cultures provided evidence for deregulation of IFN signaling in SCP32, leading to a small, but significant reduction in sensitivity to the antiproliferative effects of IFN-α and IFN-β (Fig. 5A-F). It is worth noting that a small increase in cellular fitness can cause a single SCP to outgrow others in the course of four months, as illustrated by the prediction of SCP growth dynamics in SCP mixtures propagated *in vitro*, which was solely based on the SCP growth rates that displayed minimal variation (Supplementary Fig. 3D). In addition to the cell-autonomous antitumor functions, type I IFNs can promote antitumor immune responses by activation of downstream signaling and the expression of additional cytokines (15). Concurrently, type I IFNs can induce immune evasive mechanisms that enable tumor growth (16). For example, in response to type I IFNs, tumor cells may upregulate PD-L1 and PD-L2 expression, ligands of the PD-1 immune inhibiting checkpoint, and thereby attenuate anti-tumor immunity (17–21). Thus, depending on the balance between these opposing immunomodulatory functions of type I IFNs, the tumor microenvironment may select either for or against loss of IFN signaling. Since most xenograft models for human cancer are based on immune-deficient mouse strains, it is expected that the selective pressures for tumor cell-autonomous IFN signaling are shifted and favor the expansion of cells having genetic alterations that impair IFN signaling. Additional research is required to delineate the relative contributions of cell-autonomous and non-cell autonomous factors to changes in IFN signaling during passaging of PDX models.

Supplementary Material and Methods

Cell culture and generation of SCPs

To propagate cultures or generate cell suspensions, cells were washed once with PBS buffer (Corning, Cat. No. 21-031-CV) and treated with 0.25% Trypsin EDTA (Corning, Cat. No. 25-053-CI) for approximately 10 minutes. The Z-Series Coulter Counter (Beckman) was used to determine the number of cells per mL cell suspension. SCPs were generated through single-cell cloning of the MDA-MB-468 cell line as follows: cells were plated by limited dilution into 96 well plates. Out of the 576 wells, 63 contained a single cell, as determined by visual inspection. These cells and their progeny were propagated in a 1:1 mixture of fresh and MDA-MB-468-conditioned medium for two to three weeks. Following this period, each SCP was propagated in standard growth medium.

Barcoding of SCPs, Plasmids, CRISPR/Cas9, and lentiviral transduction

Each SCP was transduced with a unique ClonTracer lentiviral vector at a transduction efficiency of approximately 10%. The lentiviruses were produced as follows: 293T cells were plated at 2.5 x 10^6 cells per 10 cm dish. The next day, 2.4 µg of the lentiviral vector of interest was mixed with 1.8 µg psPAX2 and 0.6 µg pMD2.G packaging vectors in 800 µL Opti-MEM® Reduced Serum Medium (Life Technologies, Cat. No. 31985-062). 14.4 µL of 1 mg/mL Polyethylenimine (linear; MW~25,000; PolySciences, Inc.; Cat. No. 23966) was added to this solution. The mixture was then vortexed for 20 seconds, left to incubate for 15 – 20 minutes, vortexed for 20 seconds, and then carefully transferred to the 293T culture. Medium was refreshed 16 – 24 hours post transfection and virus supernatant was collected 72 hours later, passed through a 0.45 µm Puradisc syringe filter (GE Healthcare Life Sciences, Cat. No. 6780-2504) and added to the recipient cells at different dilutions. The barcoded cells were subsequently selected by treatment with 2 µg/mL puromycin (InvivoGen, Cat. No. ant-pr-5) for at least 3 days. Out of 31 SCPs, nine did not survive selection for integration of lentivirus by treatment with puromycin and these SCPs were therefore excluded from further analysis. The identity of the barcoded SCPs was frequently verified by PCR-mediated amplification of the barcode sequences and Sanger sequencing using the following primers:

WSL_Sanger_PCR_Fw	GAACAGATTTGGAATCACACGACC
WSL_Sanger_PCR_Rv	TAAGGCCGAGTCTTATGAGCAG
WSL_Sanger_SEQ_Fw	GAAGGTGGAGAGAGAGACAGAG

The SCPs are identified by the following unique barcode sequences:

SCP01	TCTGTCACACACAGTCACTCTGAGAGACAG
SCP03	TCAGACTCTGTCACAGTGTGAGACAGAGTG
SCP05	ACTGAGACAGAGTCAGTGTGAGTGTGTCAG
SCP07	TCTGACACTCAGACTCAGTGACTGTGACTG
SCP08	AGTGACAGACACTCTGACAGACTCAGTGAG
SCP12	AGTGAGAGAGTGAGACTGTCTCAGAGAGTC
SCP13	TGTCAGAGTCTGAGTCTGTCTCACTGAGAG

SCP14	TGTGTGTCTGACAGTCTCTCTCACTGAGAG
SCP15	TCAGTCTCAGTCTCACTGTGTGTCACTCTC
SCP16	ACTGTCTGAGACAGAGAGTGTGACAGTCAG
SCP17	AGTGTCACTGTGTGACTGAGAGTCTGACAG
SCP18	ACTCAGTGAGAGACAGTGTCACAGTGAGTG
SCP22	TGACAGAGTGTGTCAGAGTGTGAGTGAGTG
SCP24	TGACTCAGAGAGTCTCTGTGTCTGTCAGAC
SCP25	TGTGTCACAGTCTCACAGAGAGACAGAGAC
SCP26	TGTGTCTGACTGACTGACAGTGACACACTG
SCP27	AGTGAGACTGAGTCAGACTGAGACTGTGAG
SCP28	AGACAGACTCTCAGTCTGTCAGACAGTGAG
SCP29	TCTGTGTCACACACTCTGTCTGAGAGTGTC
SCP31	TGAGACACTGTGACTCAGTGTGTCAGAGAG
SCP32	ACTCTGTGTCTCAGTGTGAGTGTCTGACTG
SCP33	TCTCTCTCTGAGTGTGAGACTGTGTCTCTG

SCP32 empty vector and SCP32 IFNAR2 overexpressing cell lines were generated by lentiviral transduction of non-barcoded SCP32 with the Phage-CMV-FLAG-HA-IRES-Puro and Page-CMV-hIFNAR2-FLAG-HA-IRES-Puro vectors and were kept under selection using 2 µg/mL puromycin (InvivoGen, Cat. No. ant-pr-5). SCP29 Cas9 was generated by lentiviral transduction of non-barcoded SCP29 with the LentiCas9-Blast vector (Addgene, #52962) developed by Zhang and colleagues and was kept under selection using 10 µg/mL blasticidin (InvivoGen, Cat. No. ant-bl-1) (22). The lentiviral vectors encoding for the STAT1 and non-targeting sgRNAs were generated by cloning of sgRNAs oligos (Integrated DNA Technologies) into the LentiGuide-Puro vector (Addgene, #52963) as described previously (22). The sgRNA sequences were verified by Sanger sequencing using the hU6 sequence primer GACTATCATATGCTTACCGT. The sgRNA targeting sequences used in this study are as follows:

sgNT_1	TATCGCGTAGTGCTGACGT
sgNT_2	TTACAATCGTCGGTCCAAT
sgSTAT1_1	TGCTGGCACCAGAACGAATG
sgSTAT1_3	CATGTTGTACCAAAGGATGG

Mouse experiments

On the day of transplantation, cells were collected by treatment with 0.25% trypsin (Corning, Cat. No. 25-053-CI) for approximately 10 minutes. The Z-Series Coulter Counter (Beckman) was used to determine the number of cells per mL cell suspension. In case of SCP mixtures, SCPs were mixed in equal proportions. Next, 2×10^6 cells were injected orthotopically into the fourth mammary gland of anesthetized mice. The mice were subsequently monitored twice weekly, and tumors were measured every two weeks after they were palpable. The mice were euthanized after 3 weeks, 2 months and 4 months, or when tumors reached the maximum allowed size.

Copy number analysis

Genomic DNA was isolated from one 10 cm dish per SCP and the parental cell line using the QIAamp DNA Mini Kit (Qiagen, Cat. No. 51306). The DNA concentration was measured on the Epoch microplate spectrophotometer (BioTek). Copy number was inferred from sparse shotgun sequencing data of TruSeq indexed Illumina sequencing libraries. 1ug of genomic DNA was sonicated using a Covaris instrument to +/- 300 bps. Sonicated DNA was then end-repaired/a-tailed and ligated to TruSeq Illumina adaptors using standard protocols. Adaptor ligated DNA molecules were then enriched by PCR, quantification, and pooled to achieve an average of roughly 1-2 million reads per sample, sufficient to call copy number alterations at a resolution of roughly 300kbs. The genome was partitioned into 5K bins of variable sizes based on the unique mappability of sequences across the human genome, with each bin containing the same number of mappable positions using published methods (23). To identify unique and shared copy number alterations in the clones, segment copy number values were classified as follows: amplified (greater than or equal to 3.5 copies), gain (2.5-3.5 copies), loss (1-1.5 copies) and deletion (<0.5 copies). Segments were mapped to entrez gene identifiers using the GenomicRanges and TxDb.Hsapiens.UCSC.hg19.knownGene packages. The copy number estimates for *EGFR* were derived at a higher resolution (50K bins).

Whole-exome sequencing

Six SCPs were selected for whole-exome sequencing analysis based on the observed differences in the ability of these SCPs to form colonies in soft agar and orthotopic tumors in NOD/Scid mice. Genomic DNA was isolated from one 10 cm dish per SCP and the parental cell line using the QIAamp DNA Mini Kit (Qiagen, Cat. No. 51306). The DNA concentration was measured on the Epoch microplate spectrophotometer (BioTek). The whole-exome sequence libraries were generated by the Biopolymers Facility at Harvard Medical School using the SureSelect Exome V6+COSMIC Capture Library and SureSelectXT Reagent kits (Agilent Technologies, Cat. No. 5190-9307 and G9611A). The quality of these libraries was assessed by a 2200 TapeStation analyzer using D1000 screen tapes (Agilent Technologies), and q-PCR. All samples were pooled into a single library and subjected to paired-end sequencing using two full chips and a HiSeq2500 (Illumina). Adapter sequence was trimmed using Cutadapt 1.8 and trimmed reads were mapped with bwa-0.7.8 to the GRCh37 version of the human genome. PCR duplicates were marked with Picard v.2.8.0. Mutation and indel calling was performed using the Genome Analysis Toolkit (gatk-3.3-0) with the following steps (non-default parameters are in parentheses): RealignerTargetCreator (1000G 1000G_phase1.indels.b37.vcf and Mills_and_1000G_gold_standard.indels.b37.sites.vcf from gatk), IndelRealigner , BaseRecalibrator (dbsnp_137.b37.vcf from gatk), HaplotypeCaller (without downsampling, stand_call_conf =30 and stand_emit_conf=10.0), and GenotypeGVCFs. Genetic alterations were annotated with ANNOVAR. These genetic alterations, referred to as SNVs in this study, were subsequently filtered for having passed the VariantFiltering step, and a genotype call for all samples. This led to the identification of 10,057 SNVs, 9,205 (91.53%) of which were shared by all samples. A high confidence list of 215 SNVs was generated by additional filters: (a) each of the samples had to have at least 20X coverage for the corresponding genomic loci and (b) at least one SCP had to have no variant calls AND one SCP had to have a VAF of at least 20%.

Bulk RNA-sequencing

Library preparation and RNA-sequencing of *in vitro* samples: each cell line was plated at 2 x 10^6 cells per 10 cm dish and left to propagate for 48 hours. This process was repeated at least once before collecting RNA samples through direct lysis on the 10 cm dishes using the PureLink^{TM} RNA Mini Kit (Life Technologies, Cat. No. 12183025). The RNA concentration was measured using the Epoch microplate spectrophotometer (BioTek). Prior to library generation for RNA-sequencing, the quality of the RNA was determined using a 2200 TapeStation and RNA screen tapes (Agilent Technologies). The mRNA libraries were generated by the Biopolymers Facility at Harvard Medical School using the KAPA mRNA HyperPrep kit (Roche) and included poly-A enrichment. The quality of these libraries was assessed by a 2200 TapeStation analyzer using D1000 screen tapes (Agilent Technologies), and q-PCR. In total, 34 mRNA libraries (2 samples for parental and SCP32, and 1 sample of each other SCP) were multiplexed into two pools and subsequently sequenced (paired-end, 50 cycles) using two NextSeq Mid Output flow cells per library (Illumina).

Library preparation and RNA-sequencing of tumor samples: one half of the tumors was used for the RNA-sequencing analysis and the other half for the TMT-MS analysis (see below). RNA lysates were prepared using the PureLink^{TM} RNA Mini Kit (Life Technologies, Cat. No. 12183025). The RNA concentration was measured using the Epoch microplate spectrophotometer (BioTek). Prior to library generation for RNA-sequencing, the quality of the RNA was determined using a 2200 TapeStation and RNA screen tapes (Agilent Technologies). The mRNA libraries were generated by the Bauer core facility at Harvard University using the KAPA Stranded RNA Hyperprep kit (Roche) and included polyA enrichment. The quality of these libraries was assessed by a 2200 TapeStation analyzer using D1000 screen tapes (Agilent Technologies), and q-PCR. In total, 12 mRNA libraries (3 samples per SCP) were multiplexed into a single pool and subsequently sequenced (paired-end, 75 cycles) using one NextSeq High Output flow cell (Illumina).

Adapter sequence was trimmed from raw RNA-seq reads using *Cutadapt v1.8* and the trimmed reads were mapped with hisat2-2.0.3 to the GRCh37 version of the human genome. *Rsubread v. 1.32.4*was used to quantify raw counts per gene. The threshold for expression detection was > 1 logCPM in at least 2 samples. Differential expression was calculated using the *glmFit* and *glmLRT* implementation of generalized linear model (glm) methods in *EdgeR v.3.24.3*. Hierarchical clustering using Spearman correlation and average linkage. Heatmaps were generated with the *heatmap.2* function in *gplots v3.0.3*. All analyses were performed using R.3.5.1.

Gene set enrichment analyses (GSEA) and gene signatures

Genes were ranked using the equation: -log10(FDR) * sign(logFC)) where FDR is the false discovery rate corrected p-value and logFC is the log2 fold-change calculated in *edgeR*. Gene set enrichment analysis (GSEA) was performed with *fgsea v1.8.0* using the 'Hallmark', 'Chemical and Genetic Perturbations' and 'Oncogenic' gene set collections from msigdb v6.2. Gene sets were combined prior to analysis to ensure that the correction for false discovery rate correction was applied uniformly.

In addition to GSEA, an established proliferation (24) and a custom apoptosis gene signature were used to characterize SCP tumors. The apoptosis signature was defined as the

genes that were found in at least three of the following apoptosis gene sets: DEBIASI_APOPTOSIS_BY_REOVIRUS_INFECTION_UP (25), HOLLMANN_APOPTOSIS_VIA_CD40_UP (26), HAMAI_APOPTOSIS_VIA_TRAIL_UP (27), CONCANNON_APOPTOSIS_BY_EPOXOMICIN_UP (28), GRAESSMANN_APOPTOSIS_BY_DOXORUBICIN_UP and GRAESSMANN_APOPTOSIS_BY_SERUM_DEPRIVATION_UP (29). The resulting signature included: CREM, GADD45A, IFI16, STAT1, AGTPBP1, ALDH2, ATF3, BCL2L2, BST2, CASP7, CD55, CEBPB, CEBPG, CLTB, CXCL1, CXCL10, CXCL2, DDIT3, FAS, HSPH1, IFI35, IFI44, IFIH1, IFIT1, IFIT3, IGFBP4, IRF7, ISG15, MAK16, MEST, MMD, MX2, NMI, OAS2, OASL, PHLDA2, PLK2, PPP1R15A, RPE, RSAD2, SP100, SP110, TBC1D1, TRIM21, UBE2V2, ZFP36, ZNF43, and FASL, the latter of which was added manually. The proliferation and apoptosis scores represent the average expression (TPM) of the genes in the corresponding gene sets. PDX tumors of human breast cancer (30,31) were examined for expression of IFN (32) and T-cell metagene sets (33). The IFN and T cell metagene scores were determined as follows: for each of the genes in a specific gene set, z-scores were computed across the samples using the TPM gene expression values. These z-scores were then summarized by the mean.

Single-cell RNA-sequencing

The parental MDA-MB-468 cell line was plated at 2 x 10^6 cells per 10 cm dish and left to propagate for 48 hours. This process was repeated one additional time. To generate the single cell suspension, cells were washed with 1X PBS and treated with 0.25% Trypsin EDTA for 5 minutes. The cells were then washed 3 times with 1X PBS + 0.04% BSA and resuspended in 1X PBS + 0.04% BSA at approximately 1 x 10^6 cells/mL. Single cell capturing and cDNA library generation were performed by the Bauer core facility at Harvard University using the 10X Chromium 3' library construction kit v2 following the manufacturer's instruction. The library was sequenced (paired-end, 150 cycles) using two NextSeq High Output flow cells (Illumina).

Sequencing reads were converted to fastq files using bcl2fastq v.2.20.0.422. Reads from two sequencing runs of the same sample were merged, aligned, and counted with cellranger v3.0.2 using refdata-cellranger-hg19-3.0.0 provided by 10x Genomics, Inc. All subsequent data processing and analysis was performed in *Seurat v.3.1.5*. Genes detected in <10% of cells were removed. Cells with <500 or >8000 unique gene counts, >10% mitochondrial genes, or <1000 unique unique molecular identifiers (UMI) were excluded. The Adaptively-Thresholded Low Rank Approximation (ALRA) method was used to impute missing values. Data were normalized with a scale factor of 10000 and scaled (linear model), regressing out mitochondrial content, proliferation and number of UMI. Clustering and tSNE plots were generated using principal component analysis (PCA) with 3000 variable genes ('vst' method) and 20 PCs.

Correlation of scRNA-seq data with bulk RNA-seq of each SCP was performed using 925 genes. These genes are the intersection between the top 500 genes that distinguish each SCP from parental (fold-change in TPM) from bulk RNA-seq and the top 3000 most variable genes in scRNA-seq ('vst' method).

Population doubling time

Each cell line was seeded at 1500 cells per well in six wells of a 384 well plate. The next day, live-cell imaging was performed using an IncuCyte system (Essen BioScience). Images of each well were taken at 10X magnification every 4 hours for the duration of 6 - 7 days. The IncuCyte software was used to determine well confluence for each of the acquired images. The population doubling times were computed from confluence measurements within the exponential growth phase.

Soft agar assay

Soft agar assays were performed on 6-well plates in triplicate. For each well, $2x10^4$ cells were mixed thoroughly in standard growth medium containing 0.5% agarose with low melting point (Sigma, Cat. No. A9414). These cell suspensions were transferred to wells with solidified agarose gels, i.e., 0.8% agarose in standard growth medium. Each cell suspension was allowed to solidify and subsequently covered in 1 ml standard growth medium, which was refreshed once a week. The cells were allowed to propagate for two weeks. Next, colonies were stained with 0.05% (wt/vol) iodonitrotetrazolium chloride (Sigma Aldrich, Cat. No. I10406-5G), each plate was scanned at 1200 dpi using a flatbed scanner, and ImageJ was used to count the number of colonies.

DNA barcode analysis

The in vitro SCP mix cultures were sampled each time they were passaged, i.e., approximately $2x10^6$ cells were collected following cell counting by centrifugation and stored at -80°C. Tumor and lung tissue samples were also stored at -80°C prior to genomic DNA isolation. Tumors were mechanically homogenized on ice and 50 – 100 mg tumor tissue was transferred to an Eppendorf tube for genomic DNA isolation. While lung samples were lysed completely, an aliquot equivalent to 100 mg tissue was used to isolate the genomic DNA. The genomic DNA was isolated using the QIAamp DNA Mini Kit (Qiagen, Cat. No. 51306) and DNA concentrations were measured on the Epoch microplate spectrophotometer (BioTek). The barcodes were subsequently isolated by PCR on 3 µg genomic DNA per sample (max. 1 µg per PCR) using a common forward primer and set of reverse primers with unique index sequences that allow for multiplexing of samples:

> WSL_PCR_Forward:
> AATGATACGGCGACCACCGAGATCTACACACTGACTGCAGTCTGAGTCTGACAG

> WSL_PCR_Reverse:
> CAAGCAGAAGACGGCATACGAGAT<u>NNNNNNNNNNN</u>GTGACTGGAGTTCAGACGTGT
> GCTCTTCCGATCTCTAGCACTAGCATAGAGTGCGTAGCT

Per sample, PCR products were combined and isolated using the QiaQuick PCR purification kit (Qiagen, Cat. No. 28106). The quality and concentration of the PCR products was determined using a 2200 TapeStation and D1000 screen tapes (Agilent Technologies). To generate the NGS libraries, samples were mixed in equal proportions. The libraries were isolated from a 2% agarose gel using the QiaQuick gel extraction kit (Qiagen, Cat. No. 28706) and their quality and

concentration was determined using a 2200 TapeStation and D1000 screen tapes (Agilent Technologies), and q-PCR. Each library was sequenced on a MiSeq (Illumina) by the Biopolymers Facility at Harvard Medical School using the following primers:

WSL_NGS_Barcode_Seq:
GCGACCACCGAGATCTACACACTGACTGCAGTCTGAGTCTGACAG

WSL_NGS_Index_Seq:
GATCGGAAGAGCACACGTCTGAACTCCAGTCAC

Index sequences were used to demultiplex the samples. Barcode sequences that matched the 15xWS design, a phred quality score of 10 or greater for each position and an average phred quality score greater than 30, were selected for further analysis. First, we selected the barcode sequences that were used in the experiment and normalized the barcode counts to their mean fractions in the T=0 reference samples. Next, for each sample, the barcode fractions were computed by dividing the number of reads of a given barcode by the total number of reads for all barcodes in that sample.

Mass cytometry
Mass cytometry (cytometric time-of-flight; CyTOF) was performed on 19 SCPs and the parental cell line according to the standards of the field (34). The advantage of this approach is that CyTOF is not affected by differences in SCP autofluorescence and variable expression of RFP, which is encoded by the ClonTracer DNA barcode constructs that were introduced by lentiviral transduction. In brief, cells were washed with PBS (Corning, Cat. No. 21-031-CV), trypsinized and washed with PBS containing 2% BSA (cell staining medium; "CSM"). Cells were then stained in 500 μL 1:1000 cisplatin solution (Fluidigm 201064) as a viability dye for 5 minutes and quenched with 5 mL CSM. Cells were fixed in 1.6% PFA for 10 minutes and then stored in liquid nitrogen until they could be processed further. Samples were later thawed and barcoded with unique combinations of palladium isotopes (Fluidigm 201060). Samples were then pooled and treated with DNAse I (Stemcell 07900) for 15 minutes. Cells were washed and resuspended in CSM. Antibodies to CD24 (Biolegend 311102) and CD44 (Biolegend 103002) which had been conjugated to 158Gd and 115In, respectively, were then added to the cell solution and allowed to incubate for 30 minutes. Cells were washed and fixed in 4% PFA for 30 minutes. Cells were washed in CSM again and resuspended in a 1.6% PFA solution containing 1:5000 Iridium intercalator (Fluidigm 201192) to stain DNA, thereby allowing exclusion of cell clumps and debris in downstream analyses. The sample was left overnight at four degrees Celsius. After incubation, cells were washed with distilled water 3 times, EQ four element calibration beads (Fluidigm 201078) were added to the sample, and the sample was run on the Helios CyTOF instrument. After acquisition, the data were normalized and debarcoded into fcs files containing each of the clones individually. These files were then imported into the Flowjo software. Following the exclusion of cell debris, cell clumps, dead cells, and calibration bead-cell clumps, biaxial plots comparing the expression of CD24 and CD44 were generated.

Proteomic Mass Spectrometry, Data Processing and Analysis

Sample Preparation: tumor samples were snap-frozen in liquid nitrogen. Samples were homogenized in 8 M Urea, 200 mM EPPS, pH 8.5, with protease inhibitor, and lysed by passing through a 21-gauge needle with syringe. After centrifugation at 13,000 rpm at 4°C for 10min, supernatants were used for further analysis. BCA assay was performed to determine protein concentration of each sample. Samples were reduced in 5 mM TCEP for 15min, alkylated with 10 mM iodoacetamide for 15min, and quenched with 15 mM DTT for 15min. One hundred μg protein was chloroform-methanol precipitated and re-suspended in 100 μL 200 mM EPPS pH 8.5. Protein was digested by Lys-C at a 1:100 protease-to-peptide ratio overnight at room temperature with gentle shaking. Trypsin was used for further digestion for 6 hours at 37°C at the same ratio with Lys-C. After digestion, 30 μL acetonitrile (ACN) was added into each sample to 30% final volume. Two hundred μg TMT reagent (Experiment 1: 126, 127N, 127C, 128N, 128C, 129N, 129C, 130N, 130C. Experiment 2: 126, 127N, 127C, 128N, 128C, 129N, 129C, 130N, 130C, 131) in 10 μL ACN was added to each sample. After 1 hour of labeling, 2 μL of each sample was combined, desalted, and analyzed using mass spectrometry. Total intensities were determined in each channel to calculate normalization factors. After quenching using 0.3% hydroxylamine, eleven samples were combined in a 1:1 ratio of peptides based on normalization factors. The mixture was desalted by solid-phase extraction and fractionated with basic pH reversed phase (BPRP) high performance liquid chromatography (HPLC), collected onto a 96 six well plate and combined for 24 fractions in total. Twelve fractions were desalted and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (35).

Liquid chromatography and tandem mass spectrometry method experiment 1: mass spectrometric data were collected on an Orbitrap Fusion mass spectrometer coupled to a Proxeon NanoLC-1200 UHPLC. The 100 μm capillary column was packed with 35 cm of Accucore 50 resin (2.6 μm, 150Å; ThermoFisher Scientific). Peptides were separated using a 150min gradient of 3-25% acetonitrile in 0.1% formic acid with a flow rate of ~550 μL/min. The scan sequence began with an MS1 spectrum (Orbitrap analysis, resolution 120,000, 350−1400 Th, automatic gain control (AGC) target 4E5, max. injection time 50 ms). SPS-MS3 analysis was used to reduce ion interference (36,37). The top ten precursors were then selected for MS2/MS3 analysis. MS2 analysis consisted of collision-induced dissociation (CID), quadrupole ion trap analysis, automatic gain control (AGC) 15,000, NCE (normalized collision energy) 35, q-value 0.25, maximum injection time 120 ms), and isolation window at 0.7. Following acquisition of each MS2 spectrum, we collected an MS3 spectrum in which multiple MS2 fragment ions are captured in the MS3 precursor population using isolation waveforms with multiple frequency notches. MS3 precursors were fragmented by HCD and analyzed using the Orbitrap (NCE 65, AGC 100,000, max. injection time 150 ms, resolution was 50,000 at 400 Th).

Liquid chromatography and tandem mass spectrometry method experiment 2: mass spectrometric data were collected on an Orbitrap Fusion Lumos mass spectrometer coupled to a Proxeon NanoLC-1200 UHPLC. The 100 μm capillary column was packed with 35 cm of Accucore 50 resin (2.6 μm, 150Å; ThermoFisher Scientific). Peptides were separated using a 90min gradient of 3 to 20% acetonitrile in 0.1% formic acid with a flow rate of ~520 μL/min. The scan sequence began with an MS1 spectrum (Orbitrap analysis, resolution 120,000, 350−1400

Th, automatic gain control (AGC) target 4E5, maximum injection time 50 ms). SPS-MS3 analysis was used to reduce ion interference (36,37). MS2 analysis consisted of collision-induced dissociation (CID), quadrupole ion trap analysis, automatic gain control (AGC) 2,000, NCE (normalized collision energy) 35, q-value 0.25, maximum injection time 120 ms), and isolation window at 0.7. Following acquisition of each MS2 spectrum, we collected an MS3 spectrum in which multiple MS2 fragment ions are captured in the MS3 precursor population using isolation waveforms with multiple frequency notches. MS3 precursors were fragmented by HCD and analyzed using the Orbitrap (NCE 65, AGC 150,000, maximum injection time 120 ms, resolution was 50,000 at 400 Th). This data acquisition includes high-field asymmetric-waveform ion-mobility spectrometry (FAIMS). The dispersion voltage (DV) for FAIMS was set at 5000V, the compensation voltages (CVs) were set at -40V, -60V, and -80V, and TopSpeed parameter was set at 1 sec per CV (38).

Data analysis: mass spectra were processed using a Sequest-based pipeline (39,40). Spectra were converted to mzXML using a modified version of ReAdW.exe. Database searching included all entries from the Human UniProt database (downloaded: 2014-02-04). This database was concatenated with one composed of all protein sequences in the reversed order. Searches were performed using a 50-ppm precursor ion tolerance for total protein level analysis. The product ion tolerance was set to 0.9 Da. TMT tags on lysine residues and peptide N termini (+229.163 Da) and carbamidomethylation of cysteine residues (+57.021 Da) were set as static modifications, while oxidation of methionine residues (+15.995 Da) was set as a variable modification. Peptide-spectrum matches (PSMs) were adjusted to a 1% false discovery rate (FDR) (41,42). PSM filtering was performed using a linear discriminant analysis (LDA), as described previously (40), while considering the following parameters: XCorr, ΔCn, missed cleavages, peptide length, charge state, and precursor mass accuracy. For TMT-based reporter ion quantitation, we extracted the summed signal-to-noise (S:N) ratio for each TMT channel and found the closest matching centroid to the expected mass of the TMT reporter ion. For protein-level comparisons, PSMs were identified, quantified, and collapsed to a 1% peptide false discovery rate (FDR) and then collapsed further to a final protein-level FDR of 1%, which resulted in a final peptide level FDR of <0.1%. Moreover, protein assembly was guided by principles of parsimony to produce the smallest set of proteins necessary to account for all observed peptides. Proteins were quantified by summing reporter ion counts across all matching PSMs, as described previously (40). PSMs with poor quality, MS3 spectra with more than ten TMT reporter ion channels missing, MS3 spectra with TMT reporter summed signal-to-noise of less than 100 or having no MS3 spectra were excluded from quantification (43). Each reporter ion channel was summed across all quantified proteins and normalized assuming equal protein loading of all 10 samples. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with dataset identifier PXD022325 (TMT1: experiment 1, TMT2: experiment 2).

ELISA for interferon-β
At the time of tumor collection, tumor fragments were snap-frozen using liquid nitrogen and stored at -80°C. Protein lysates were generated using 500 μL chilled RIPA buffer per tumor fragment and by applying manual disruption. The BCA assay (Thermo Scientific, Cat. No.

23228) was used to quantify the protein concentrations. The samples were then normalized to 1 µg/µL with RIPA. The concentration of human interferon-β in the sample diluents was determined using the VeriKine™ Human Interferon Beta ELISA Kit (PBL Assay Science, Cat. No. 41410-1, Lot. No. 7419).

Dose-response curve assays
Recombinant human and mouse interferon-alpha A (IFN-α, R&D Systems, Cat. No. 11100-1 and 12100-1), human and mouse interferon-beta (IFN-β, R&D Systems, Cat. No. 8499-IF-010 and 8234-MB), and human and mouse interferon-gamma (IFN-γ, R&D Systems, Cat. No. 285-IF-100 and 485-MI-100) were diluted to 40 - 100 µg/mL in a PBS buffer containing Tween-20 at a final concentration of 0.3%. Each cell line was seeded at 1200 cells per well in 384 well plates in a total volume of 50 µL culture media composed of DMEM + 2% heat-inactivated FBS + 5 units/mL penicillin and 5 µg/mL streptomycin. The next day, IFNs were added using the HP D300e digital dispenser (HP, Cat. No. D300e). Cells were then left to propagate for 5 to 7 days and subsequently fixed and stained by a 30-minute incubation in PBS containing 3.6% formaldehyde (Sigma, Cat. No. 252549) and 2 µg/mL Hoechst 33342 (Life Technologies, Cat. No. H1399). Following a final PBS wash, the wells were imaged using the Acumen eX3 high-content imager (TTP LabTech). The Cellista software (TTP LabTech) was used for automated image analysis to estimate the number of cells per well. Following normalization to vehicle control (PBS + 0.3% Tween-20), dose-response curves were fitted using the software Prism 8 (GraphPad) and the built-in four parameter variable slope equation.

Quantitative PCR
Cell lines were plated at 2 x 10^6 cells per 10 cm dish and left to propagate for 48 hours prior to collection of RNA samples through direct lysis on the 10 cm dishes using the PureLinkTM RNA Mini Kit (Life Technologies, Cat. No. 12183025). The RNA concentration was measured using the Epoch microplate spectrophotometer (BioTek). The cDNA was generated from equal amounts of RNA for each sample using the Quanta Qscript cDNA synthesis kit (Quanta Biosciences, Cat. No. 101414-098). Each q-PCR included a standard curve, which was used to compute PCR efficiency. For each sample, the expression of the gene of interest was normalized to the expression of housekeeping gene RPL13 or UBC. The q-PCR primers used in this study include:

RPL13_Fw	GAG ACA GTT CTG CTG AAG AAC TGA A
RPL13_Rv	TCC GGA CGG GCA TGA C
UBC_Fw	CTG GAA GAT GGT CGT ACC CTG
UBC_Rv	GGT CTT GCC AGT GAG TGT CT
IFNAR2_Fw	TCA TGG TGT ATA TCA GCC TCG T
IFNAR2_Rv	AGT TGG TAC AAT GGA GTG GTT TT
STAT1_Fw	TGA AGA TTA CGC TTG CTT TTC CT
STAT1_Rv	CAG CTT GAC TCA AAA TTC CTG GA

Supplementary figures

Supplementary Fig. 1. Genomic and transcriptional heterogeneity among SCPs derived from a single TNBC cell line.
a-b, Colony formation assay for MDA-MB-468 using regular growth media or a 1:1 mixture of regular growth media and media collected from a semi-confluent MDA-MB-468 culture (conditioned culture media). **a**, Representative image of one out of three independent colony formation assays. **b**, Bar chart showing normalized number of colonies in soft agar. Data of three independent experiments summarized by mean ± SEM. Unpaired t-test was used to determine statistical significance. **c**, Heatmap showing copy number estimates per gene across the whole genome. Cell lines were clustered based on Spearman correlation and average linkage. **d**, Bar chart showing *EGFR* copy number estimates for all SCPs and the parental cell line. Data based on binned (50K) copy number data. **e-f**, Correlation matrix for copy number (**e**) and gene expression (**f**) data of the SCPs and parental cell line, and published copy number and gene expression data for other TNBC cell lines. Color scale: Spearman's ρ-value. **g**, Bar chart showing the relative *EGFR* mRNA expression levels in all SCPs normalized to the parental cell line. **h**, Dot plot showing linear regression analysis for *EGFR* copy number estimates (x-axis) and *EGFR* mRNA expression (y-axis) in the SCPs.
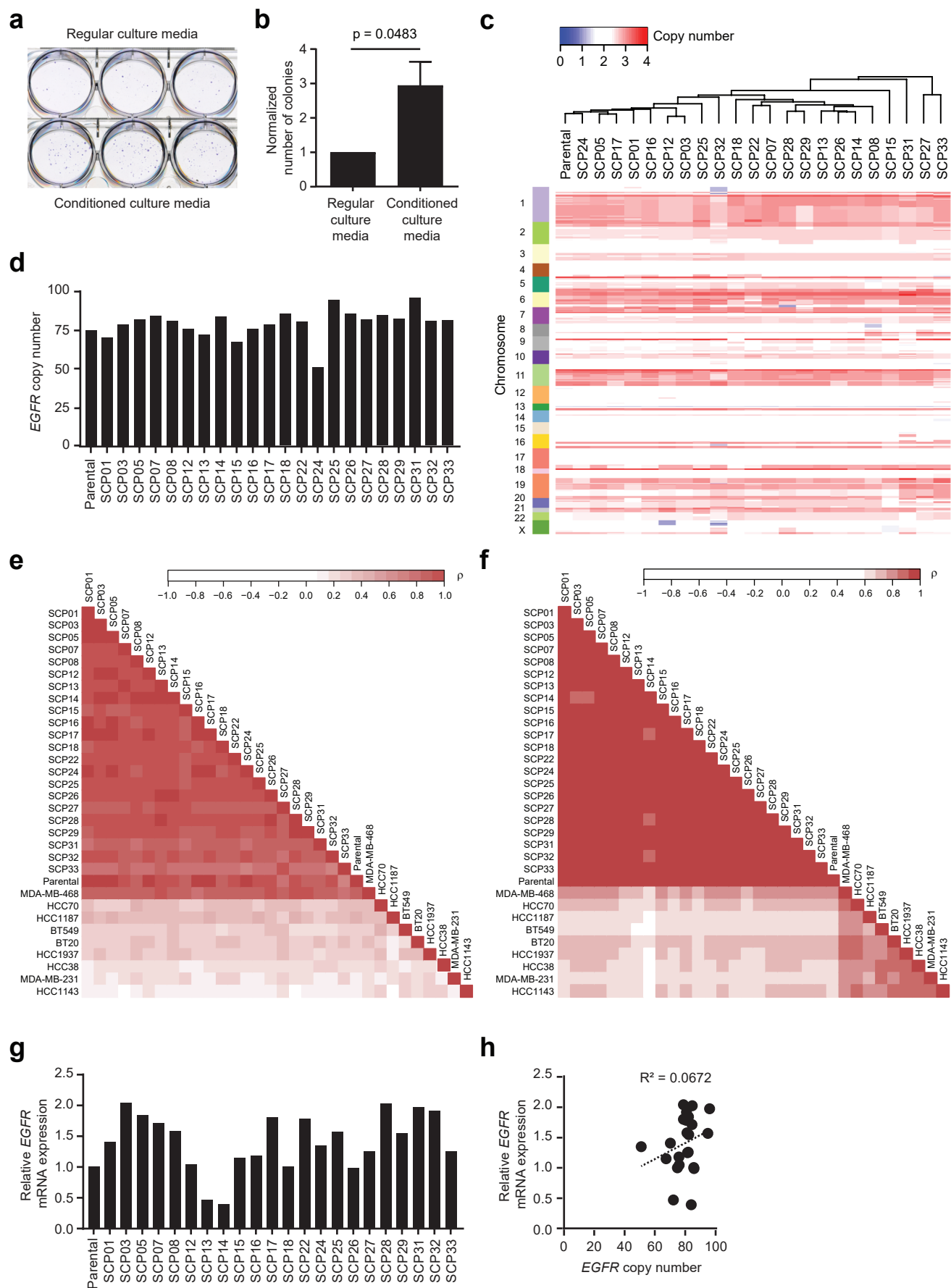
Supplementary Figure 1

Supplementary Fig. 2. SCPs display minimal variation in morphology and are markedly different with regard to their ability to form colonies in soft agar.
a, Representative phase-contrast images of all SCPs and the parental culture taken at 10x magnification. Scale bar represents 100 µm. **b**, Dot plot showing linear regression analysis for SCP population doubling time in hours (x-axis) and number of colonies formed by SCPs in soft agar (y-axis). Data shown as mean of three independent experiments. **c**, Dot plot showing number of colonies formed by individual SCPs and the parental cell line. Welch's t-test was used to determine statistical significance between the two major transcriptional subgroups.

a

Parental

SCP01

SCP03

SCP05

SCP07

SCP08

SCP12

SCP13

SCP14

SCP15

SCP16

SCP17

SCP18

SCP22

SCP24

SCP25

SCP26

SCP27

SCP28

SCP29

SCP31

SCP32

SCP33

b

R^2 = 0.0043

Number of colonies

Population doubling time (hours)

c

p = 0.9828

Number of colonies

Group 1   Group 2
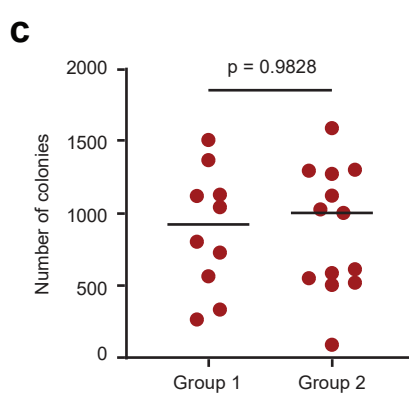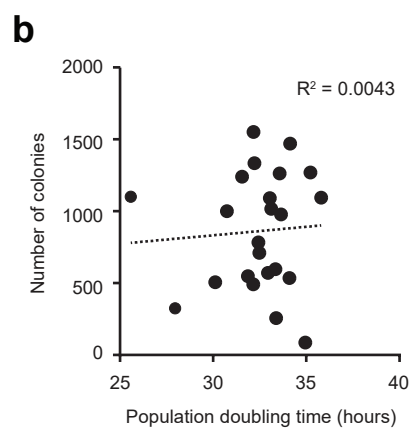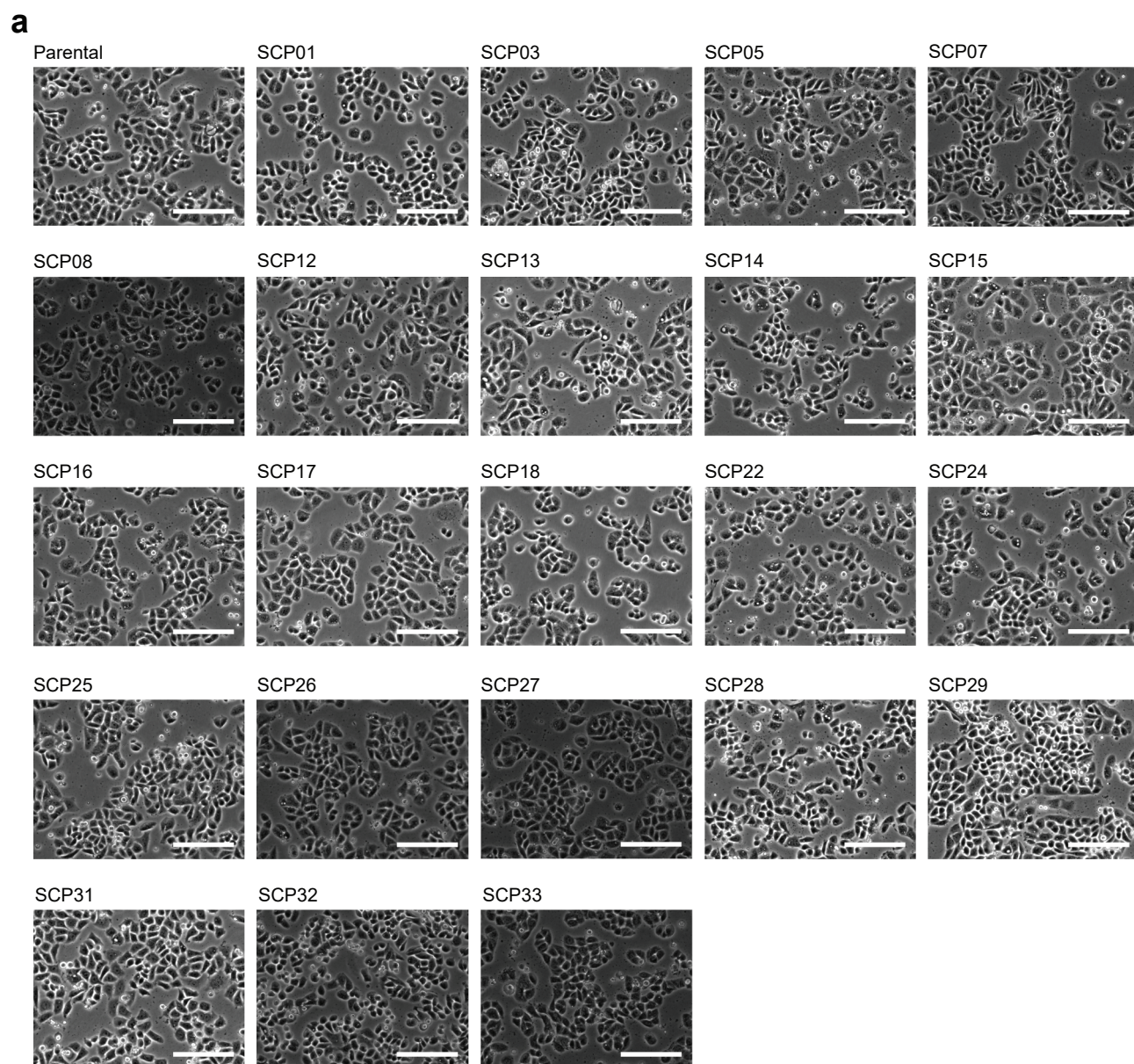
Supplementary Figure 2

Supplementary Fig. 3. Barcode analysis of tumors derived from SCP mixtures reveal divergent growth dynamics of SCP mixtures during tumor development and *in vitro* culture.

a, Dot plot showing the mass of tumors collected at four months post orthotopic transplantation of the parental cell line (n = 4) and a SCP mixture in NOD/Scid mice (n = 6). **b**, Line chart showing the change in tumor volume over time for tumors derived from the parental cell line or a SCP mixture (n = 5 each). **c**, Stacked bar chart showing barcode distributions in tumors of a SCP mixture collected at indicated time points post orthotopic transplantation (n = 4 - 7). An independent repeat experiment of Figure 3B. **d-f**, A SCP mixture was divided over five independent cultures, and each was propagated *in vitro* as monolayer in DMEM + 10% FBS for four months. **d**, Dot plot showing the population doubling time (PDT) of the parental cell line (n=3, data derived from Figure 2B) and five independent cultures of a SCP mixture (n = 36). Data shown as mean ± SD. **e**, Dot plot showing the number of doublings per day calculated at each passage of the five independent cultures of the SCP mixture in **d** (n = 5). Data shown as mean. Dotted line: linear regression analysis. **f**, Stacked bar chart showing barcode distributions in samples collected from five independent cultures of a SCP mixture at indicated time points (n = 4 - 5). One of the independent cultures was discarded between the two- and four-month time points because of contamination. **g**, Heatmap showing a correlation matrix for the barcode distributions shown in **c** and **f**. Color scale: Spearman's ρ-value. Asterisks: ***, p < 0.001; ** p < 0.01; * p < 0.05. **h**, Stacked bar chart showing predicted barcode distributions for SCP mixtures propagated as a monolayer in DMEM + 10% FBS based on the PDT of each SCP as determined in Figure 2B. In **a-b**, data shown as mean ± SEM. In **a** and **d**, Welch's t-test was used to determine statistical significance. In **c**, **f**, and **h**, colors represent different SCPs, each having a unique DNA barcode. SCPs mixtures: SCPs were mixed in equal proportions. The make-up of each SCP mixture is provided in the color legends adjacent to each stacked bar chart; a cross indicates that the SCP was excluded in the mixture. Ref: mean of reference samples collected prior to injection of the SCP mixtures.

Supplementary Figure 3

Supplementary Fig. 4. SCP mixtures cultured *in vitro* using distinct growth media formulations show divergent growth dynamics.
a, Stacked bar chart showing barcode distributions in samples collected from five independent cultures of a SCP mixture propagated as monolayer in DMEM + 10% FBS for 1 month, and subsequently in DMEM + 10% FBS or RPMI + 10% FBS for an additional 3 months (n = 3 - 5). Two of the independent cultures in RPMI + 10% FBS were discarded between the two- and three-month time points because of contamination. **b**, Growth curves for the SCP cultures propagated in **e**. **c**, Heatmap showing a correlation matrix for the barcode distributions shown in **a**. Color scale: Spearman's ρ-value. Asterisks: ***, p < 0.001; ** p < 0.01; * p < 0.05. **d**, Dot plot showing mass of tumors collected at four months post orthotopic transplantation of two SCP mixtures into NOD/Scid mice, one containing a newly barcoded version of SCP32 (n = 4 - 5). Welch's t-test was used to determine statistical significance. **e**, Stacked bar chart showing barcode distributions in tumor samples of the SCP mixture containing a newly barcoded version of SCP32 (n = 4). In **b** and **d**, data shown as mean ± SEM. In **a** and **e**, colors represent different SCPs, each having a unique DNA barcode. SCPs mixtures: SCPs were mixed in equal proportions. The make-up of each SCP mixture is provided in the color legends adjacent to each stacked bar chart; a cross indicates that the SCP was excluded in the mixture. Ref: mean of reference samples collected prior to injection of the SCP mixtures.

Supplementary Figure 4

Supplementary Fig. 5. SCPs display considerable variation in their ability to form tumors in immune-compromised mice.

The SCP cultures and the parental cell line were transplanted orthotopically into NOD/Scid mice. **a**, Tumor growth curves. Colors represent individual tumors. **b**, Stacked bar chart showing number of mice that were engrafted per cell line (n = 4 - 9). Red: tumor bearing. Black: tumor free. **c**, Dot and box plot showing tumor volumes (n = 1 - 9). One-way ANOVA and Dunnett's multiple comparison test were used to determine statistical significance compared to parental tumors. **d**, Dot plot showing linear regression analysis for SCP population doubling time in hours (x-axis) and SCP tumor mass in grams (y-axis). **e**, Dot plot showing linear regression analysis for the number of colonies formed by SCPs in soft agar (x-axis) and SCP tumor mass in grams (y-axis). **d-e**, Data shown as the mean of three independent experiments or the mean of individual tumors.

a
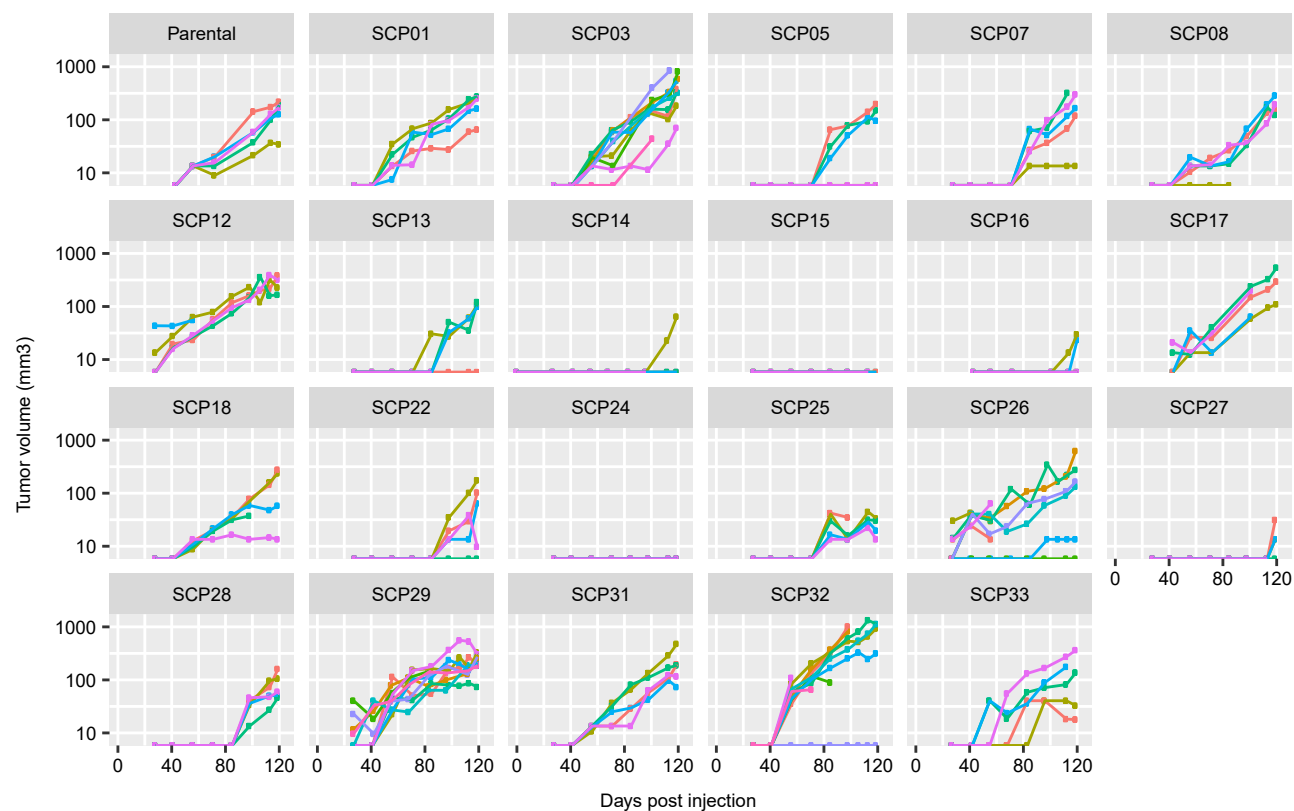
b
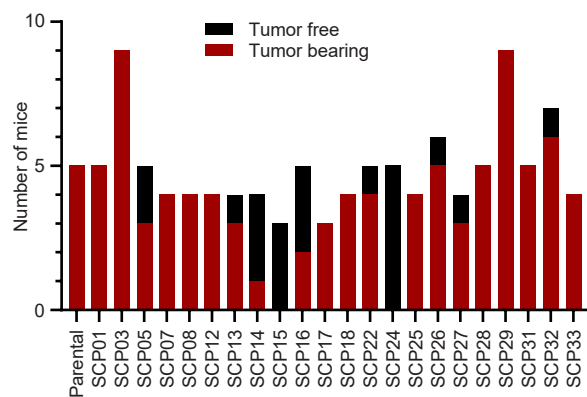
c
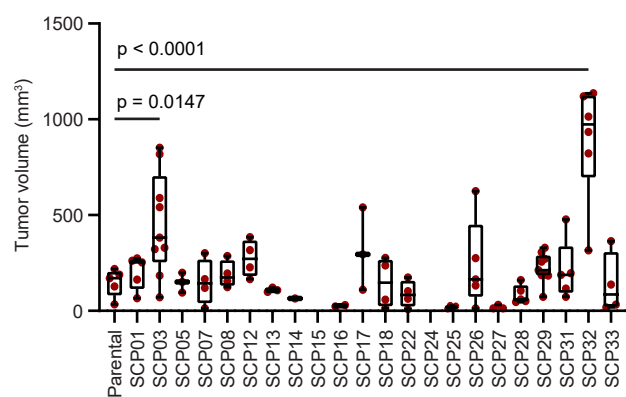
d
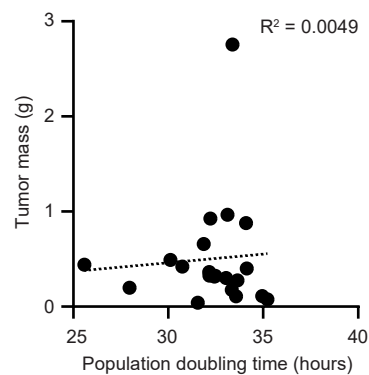
e
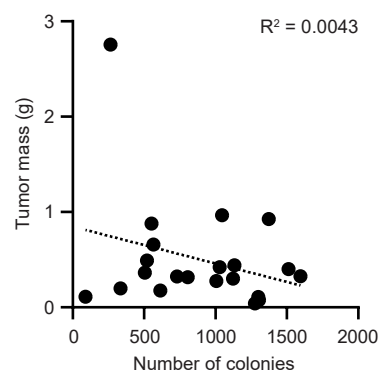
Supplementary Figure 5

Supplementary Fig. 6. SCP tumorigenic potential does not correlate with the fraction of tumor-initiating cells or EGFR expression.

a, Dot plots showing CD44 and CD24 levels in a subset of SCPs and the parental cell line. The lower right quadrant was used to determine the proportion of CD44^{high}CD24^{low} cells. Representative of two independent CyTOF experiments. **b**, Dot plot showing the proportion of CD44^{high}CD24^{low} cells (%) in a subset of SCPs and the parental cell line (n = 2 independent experiments). One-way ANOVA and Dunnett's multiple comparison test were used to determine statistical significance compared to the parental cell line. **c**, Dot plot showing linear regression analysis for the proportion of CD44^{high}CD24^{low} cells (x-axis) and SCP tumor mass in grams (y-axis). Data shown as the mean of two independent experiments or the mean of individual tumors. **d**, Dot plots showing linear regression analysis for *EGFR* mRNA expression (x-axis) and SCP tumor mass in grams (y-axis). Data shown as the mean of individual tumors.

a

Parental: 1.51 | 97.4 | 0.16 | 0.90
SCP01: 2.14 | 83.2 | 0.93 | 13.7
SCP03: 1.41 | 90.8 | 0.31 | 7.50
SCP05: 0.91 | 92.7 | 0.32 | 6.09

SCP07: 0.96 | 96.6 | 0.23 | 2.23
SCP08: 2.13 | 96.3 | 0.14 | 1.45
SCP12: 0.60 | 87.6 | 0.38 | 11.4
SCP13: 0.94 | 96.8 | 0.24 | 1.98

SCP14: 1.70 | 97.7 | 0.12 | 0.47
SCP15: 4.64 | 59.0 | 4.02 | 32.4
SCP16: 1.02 | 98.6 | 0.099 | 0.28
SCP17: 0.56 | 84.6 | 0.37 | 14.5

SCP28: 14.0 | 56.6 | 4.16 | 25.2
SCP31: 5.34 | 87.7 | 0.58 | 6.37
SCP32: 5.73 | 90.0 | 0.52 | 3.73
SCP33: 5.78 | 89.3 | 0.58 | 4.34

CD24 (y-axis), CD44 (x-axis)

b

CD44^{high} CD24^{low} (%)

p = 0.0019
p = 0.0142

Parental, SCP01, SCP03, SCP05, SCP07, SCP08, SCP12, SCP13, SCP14, SCP15, SCP16, SCP17, SCP28, SCP31, SCP32, SCP33

c

$R^2 = 1.6 \times 10^{-6}$

Tumor mass (g) vs CD44^{high} CD24^{low} (%)

d

$R^2 = 0.1056$

Mean tumor mass (g) vs *EGFR* expression (normalized to parental)

Supplementary Figure 6

Supplementary Fig. 7. SCP tumors are characterized by induction of IFN target genes with SCP32 showing the lowest activity.

a, Dot plot showing mass of tumors collected at two months post orthotopic transplantation of SCPs into NOD/Scid mice (n = 3 each). Data shown as mean ± SEM. One-way ANOVA and Dunnett's multiple comparison test were used to determine statistical significance comparing each SCP to the others. **b**, Heatmap showing 1497 differentially expressed genes (Log2 FC > 1.5 and FDR < 0.05) in SCP tumors, comparing tumors of SCP32 with tumors of other SCPs. Data shown as row z-scores of Log2-transformed TPM. Genes and samples were clustered based on Spearman correlation and average linkage. Exact test was used to determine statistical significance. **c**, Heatmap showing mRNA expression of individual genes within the apoptosis signature gene set. Data shown as row z-scores of Log2-transformed TPM. Genes were clustered based on Spearman correlation and average linkage. **d**, Heatmap showing 729 differentially expressed proteins (Log2 FC > 0.5 and FDR < 0.2) in SCP tumors, comparing tumors of SCP32 with tumors of other SCPs. Data shown as row z-scores of Log2-transformed TMT-MS data. Proteins and samples were clustered based on Spearman correlation and average linkage. Unpaired t-test and correction for multiple testing by FDR was used to determine statistical significance. **e**, Table with top five enriched GO-terms for proteins that have a lower expression in tumors of SCP32 compared to tumors of the other SCPs. **f**, Heatmap showing differential expression of the individual proteins within the GO-term Response to type I interferon. Data shown as row z-scores of Log2-transformed TMT-MS data. Proteins were clustered based on Spearman correlation and average linkage. **g-i**, Dose response curves for the parental, MDA-MB-468 cell line treated with human and mouse IFN-α (**g**), IFN-β (**h**), and IFN-γ (**i**) for 96 hours. Three independent experiments summarized by mean ± SEM.

a

Tumor mass (g)

p = 0.0071
p = 0.0723
p = 0.0180

SCP01 SCP03 SCP29 SCP32

b

z-score
-3 0 3

SCP29B_T2, SCP29B_T1, SCP29B_T3, SCP03B_T3, SCP03B_T2, SCP03B_T1, SCP01B_T3, SCP01B_T1, SCP01B_T2, SCP32B_T2, SCP32B_T1, SCP32B_T3

c

z-score
-2 0 2

FAS, ALDH2, CXCL2, CXCL1, ZFP36, CLTB, CEBPB, TBC1D1, PPP1R15A, DDIT3, CD55, UBE2V2, CREM, ZNF43, AGTPBP1, PLK2, IGFBP4, GADD45A, CXCL10, ATF3, OAS2, IFIT3, OASL, STAT1, IFIH1, RSAD2, IFIT1, IFI44, TRIM21, BCL2L2, ISG15, CEBPG, SP110, SP100, IRF7, BST2, IFI35, MX2, RPE, MEST, MAK16, CASP7, NMI, HSPH1, MMD, PHLDA2

T1 T2 T3 T1 T2 T3 T1 T2 T3 T1 T2 T3
SCP01   SCP03   SCP29   SCP32

d

z-score
-2 0 2

SCP32_T2, SCP32_T1, SCP32_T3, SCP03_T2, SCP03_T1, SCP03_T3, SCP29_T2, SCP29_T1, SCP29_T3

e

Term ID	Term description	FDR
GO:0034340	response to type I interferon	4.93E-15
GO:0060337	type I interferon signaling pathway	3.39E-14
GO:0060333	interferon-gamma-mediated signaling pathway	9.72E-10
GO:0045087	innate immune response	8.39E-09
GO:0006955	immune response	3.56E-08

f

z-score
-2 0 2

PTPN11, SETD2, UBE2K, SCRIB, TRIM56, MAVS, LSM14A, WNT5A, SHMT2, IRAK1, IRF3, MUL1, ADAR, TBK1, CNOT7, IRF6, CDC37, TTLL12, HSP90AB1, TYK2, IP6K2, MYD88, IFNAR1, GBP2, PTPN1, PTPN2, IFITM3, HLA-E, IFI35, IRF9, JAK1, SAMHD1, USP18, FADD, YTHDF2, PSMB8, STAT1, ABCE1, ISG15, HLA-F, IFIT1, BST2, OAS1, OAS2, OAS3, IFIT3, MX1, STAT2, METTL3, CACTIN, SP100, PTPN6, YTHDF3

+ + - - + + - - Barcode
T1 T2 T1 T2 T1 T2 T1 T2
SCP29   SCP32

g

Relative well confluence

human
mouse

Interferon-α A, Log$_{10}$ [IU/mL]

h

Relative well confluence

human
mouse

Interferon-β, Log$_{10}$ [ng/mL]

i

Relative well confluence

human
mouse

Interferon-γ, Log$_{10}$ [ng/mL]

Supplementary Figure 7

Supplementary Fig. 8. PDX tumors display changes in IFN signaling during engraftment and serial passaging in immune-compromised mice.
a, Line chart showing IFN metagene scores for PDX tumors collected at different passages. Colored lines represent 27 distinct PDX models for human breast cancer. Clinical sample: primary human breast tumor. P0 - P14: passage number. **b**, Scatter plot showing correlation between IFN and T cell metagene scores for 90 clinical samples.

a
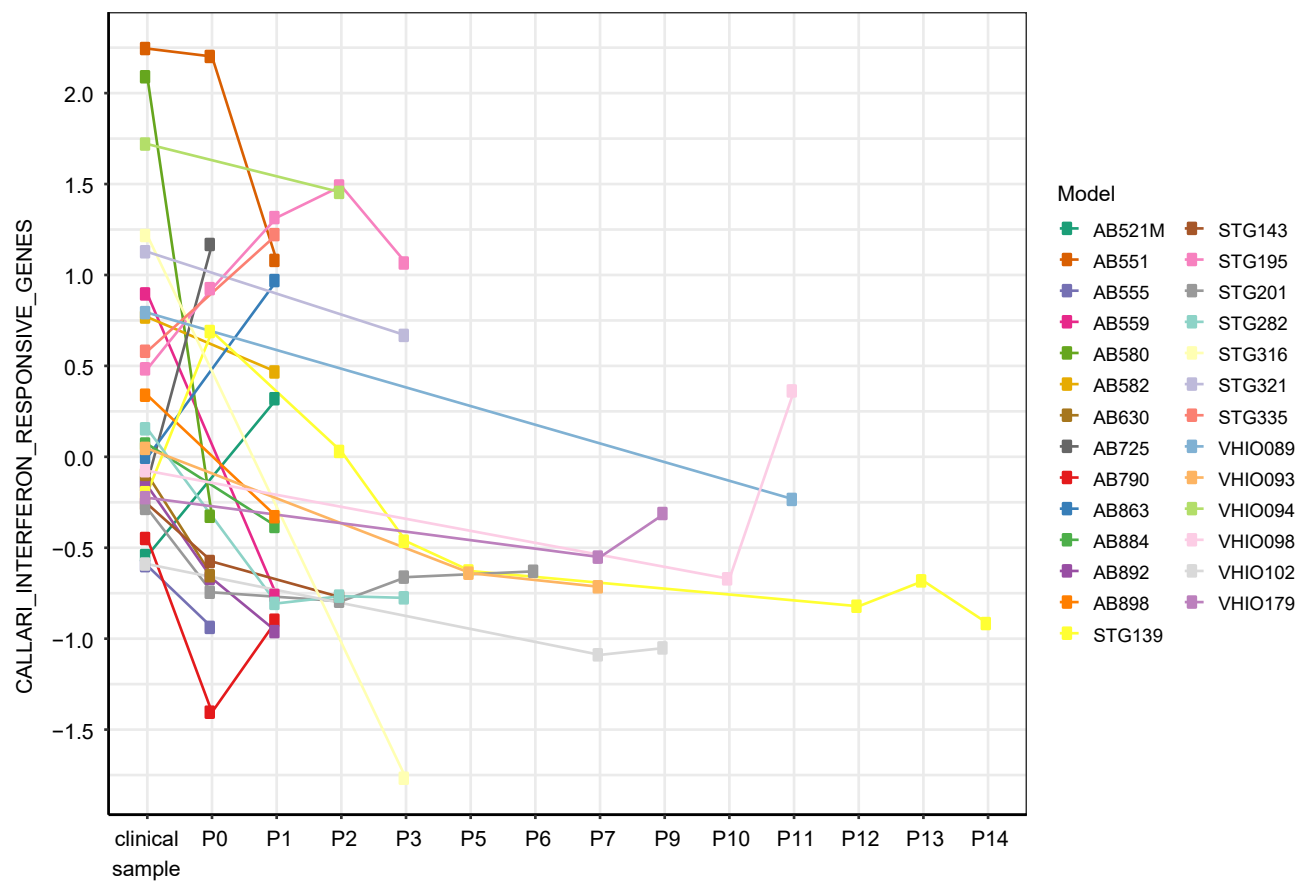
b
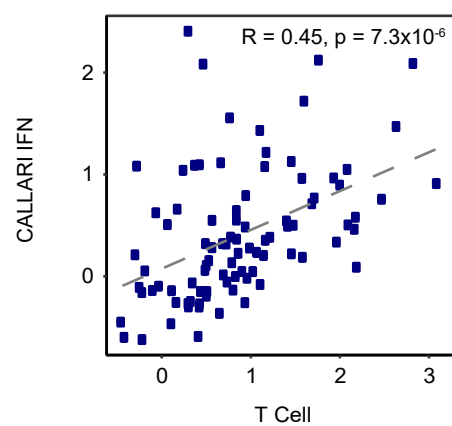
Supplementary Figure 8

Supplementary References

1. Klijn C, Durinck S, Stawiski EW, Haverty PM, Jiang Z, Liu H, et al. A comprehensive transcriptional portrait of human cancer cell lines. Nat Biotechnol. 2015 Mar;33(3):306–12.

2. Marusyk A, Tabassum DP, Altrock PM, Almendro V, Michor F, Polyak K. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. Nature. 2014 Oct 2;514(7520):54–8.

3. Naffar-Abu Amara S, Kuiken HJ, Selfors LM, Butler T, Leung ML, Leung CT, et al. Transient commensal clonal interactions can drive tumor metastasis. Nat Commun. 2020 Nov 16;11(1):5799.

4. Calbo J, van Montfort E, Proost N, van Drunen E, Beverloo HB, Meuwissen R, et al. A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer. Cancer Cell. 2011 Feb 15;19(2):244–56.

5. Cleary AS, Leonard TL, Gestl SA, Gunther EJ. Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. Nature. 2014 Apr 3;508(7494):113–7.

6. Inda M-M, Bonavia R, Mukasa A, Narita Y, Sah DWY, Vandenberg S, et al. Tumor heterogeneity is an active process maintained by a mutant EGFR-induced cytokine circuit in glioblastoma. Genes Dev. 2010 Aug 15;24(16):1731–45.

7. Vinci M, Burford A, Molinari V, Kessler K, Popov S, Clarke M, et al. Functional diversity and cooperativity between subclonal populations of pediatric glioblastoma and diffuse intrinsic pontine glioma cells. Nat Med. 2018;24(8):1204–15.

8. Wu M, Pastor-Pareja JC, Xu T. Interaction between Ras(V12) and scribbled clones induces tumour growth and invasion. Nature. 2010 Jan 28;463(7280):545–8.

9. Han J, Zhang L, Guo H, Wysham WZ, Roque DR, Willson AK, et al. Glucose promotes cell proliferation, glucose uptake and invasion in endometrial cancer cells via AMPK/mTOR/S6 and MAPK signaling. Gynecol Oncol. 2015 Sep;138(3):668–75.

10. Hou Y, Zhou M, Xie J, Chao P, Feng Q, Wu J. High glucose levels promote the proliferation of breast cancer cells through GTPases. Breast Cancer Dove Med Press. 2017;9:429–36.

11. Kerksick C, Willoughby D. The antioxidant role of glutathione and N-acetyl-cysteine supplements and exercise-induced oxidative stress. J Int Soc Sports Nutr. 2005 Dec 9;2:38–44.

12. Ortega AL, Mena S, Estrela JM. Glutathione in cancer cell death. Cancers. 2011 Mar 11;3(1):1285–310.

13. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 2003 Apr 1;100(7):3983–8.

14. Fillmore CM, Kuperwasser C. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. Breast Cancer Res BCR. 2008;10(2):R25.

15. Saleiro D, Platanias LC. Interferon signaling in cancer. Non-canonical pathways and control of intracellular immune checkpoints. Semin Immunol. 2019 Jun 1;43:101299.

16. Snell LM, McGaha TL, Brooks DG. Type I Interferon in Chronic Virus Infection and Cancer. Trends Immunol. 2017;38(8):542–57.

17. Brockwell NK, Owen KL, Zanker D, Spurling A, Rautela J, Duivenvoorden HM, et al. Neoadjuvant Interferons: Critical for Effective PD-1-Based Immunotherapy in TNBC. Cancer Immunol Res. 2017;5(10):871–84.

18. Garcia-Diaz A, Shin DS, Moreno BH, Saco J, Escuin-Ordinas H, Rodriguez GA, et al. Interferon Receptor Signaling Pathways Regulating PD-L1 and PD-L2 Expression. Cell Rep. 2017 09;19(6):1189–201.

19. Ma H, Yang W, Zhang L, Liu S, Zhao M, Zhou G, et al. Interferon-alpha promotes immunosuppression through IFNAR1/STAT1 signalling in head and neck squamous cell carcinoma. Br J Cancer. 2019;120(3):317–30.

20. Morimoto Y, Kishida T, Kotani S-I, Takayama K, Mazda O. Interferon-β signal may up-regulate PD-L1 expression through IRF9-dependent and independent pathways in lung cancer cells. Biochem Biophys Res Commun. 2018 09;507(1–4):330–6.

21. Terawaki S, Chikuma S, Shibayama S, Hayashi T, Yoshida T, Okazaki T, et al. IFN-α directly promotes programmed cell death-1 transcription and limits the duration of T cell-mediated immunity. J Immunol Baltim Md 1950. 2011 Mar 1;186(5):2772–9.

22. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods. 2014 Aug;11(8):783–4.

23. Baslan T, Kendall J, Ward B, Cox H, Leotta A, Rodgers L, et al. Optimizing sparse sequencing of single cells for highly multiplex copy number profiling. Genome Res. 2015 May;25(5):714–24.

24. Venet D, Dumont JE, Detours V. Most random gene expression signatures are significantly associated with breast cancer outcome. PLoS Comput Biol. 2011 Oct;7(10):e1002240.

25. DeBiasi RL, Clarke P, Meintzer S, Jotte R, Kleinschmidt-Demasters BK, Johnson GL, et al. Reovirus-induced alteration in expression of apoptosis and DNA repair genes with potential roles in viral pathogenesis. J Virol. 2003 Aug;77(16):8934–47.

26. Hollmann CA, Owens T, Nalbantoglu J, Hudson TJ, Sladek R. Constitutive activation of extracellular signal-regulated kinase predisposes diffuse large B-cell lymphoma cell lines to CD40-mediated cell death. Cancer Res. 2006 Apr 1;66(7):3550–7.

27. Hamaï A, Richon C, Meslin F, Faure F, Kauffmann A, Lecluse Y, et al. Imatinib enhances human melanoma cell susceptibility to TRAIL-induced cell death: Relationship to Bcl-2 family and caspase activation. Oncogene. 2006 Dec 7;25(58):7618–34.

28. Concannon CG, Koehler BF, Reimertz C, Murphy BM, Bonner C, Thurow N, et al. Apoptosis induced by proteasome inhibition in cancer cells: predominant role of the p53/PUMA pathway. Oncogene. 2007 Mar 15;26(12):1681–92.

29. Graessmann M, Berg B, Fuchs B, Klein A, Graessmann A. Chemotherapy resistance of mouse WAP-SVT/t breast cancer cells is mediated by osteopontin, inhibiting apoptosis downstream of caspase-3. Oncogene. 2007 May 3;26(20):2840–50.

30. Bruna A, Rueda OM, Greenwood W, Batra AS, Callari M, Batra RN, et al. A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. Cell. 2016 22;167(1):260-274.e22.

31. Georgopoulou D, Callari M, Rueda OM, Shea A, Martin A, Giovannetti A, et al. Landscapes of cellular phenotypic diversity in breast cancer xenografts and their impact on drug response. Nat Commun. 2021 Mar 31;12(1):1998.

32. Callari M, Musella V, Di Buduo E, Sensi M, Miodini P, Dugo M, et al. Subtype-dependent prognostic relevance of an interferon-induced pathway metagene in node-negative breast cancer. Mol Oncol. 2014 Oct;8(7):1278–89.

33. Callari M, Cappelletti V, D'Aiuto F, Musella V, Lembo A, Petel F, et al. Subtype-Specific Metagene-Based Prediction of Outcome after Neoadjuvant and Adjuvant Treatment in Breast Cancer. Clin Cancer Res Off J Am Assoc Cancer Res. 2016 Jan 15;22(2):337–45.

34. Spitzer MH, Nolan GP. Mass Cytometry: Single Cells, Many Features. Cell. 2016 May 5;165(4):780–91.

35. Navarrete-Perea J, Yu Q, Gygi SP, Paulo JA. Streamlined Tandem Mass Tag (SL-TMT) Protocol: An Efficient Strategy for Quantitative (Phospho)proteome Profiling Using Tandem Mass Tag-Synchronous Precursor Selection-MS3. J Proteome Res. 2018 01;17(6):2226–36.

36. Gygi JP, Yu Q, Navarrete-Perea J, Rad R, Gygi SP, Paulo JA. Web-Based Search Tool for Visualizing Instrument Performance Using the Triple Knockout (TKO) Proteome Standard. J Proteome Res. 2019 01;18(2):687–93.

37. Paulo JA, O'Connell JD, Gygi SP. A Triple Knockout (TKO) Proteomics Standard for Diagnosing Ion Interference in Isobaric Labeling Experiments. J Am Soc Mass Spectrom. 2016;27(10):1620–5.

38. Schweppe DK, Rusin SF, Gygi SP, Paulo JA. Optimized Workflow for Multiplexed Phosphorylation Analysis of TMT-Labeled Peptides Using High-Field Asymmetric Waveform Ion Mobility Spectrometry. J Proteome Res. 2020 03;19(1):554–60.

39. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom. 1994 Nov;5(11):976–89.

40. Huttlin EL, Jedrychowski MP, Elias JE, Goswami T, Rad R, Beausoleil SA, et al. A tissue-specific atlas of mouse protein phosphorylation and expression. Cell. 2010 Dec 23;143(7):1174–89.

41. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods. 2007 Mar;4(3):207–14.

42. Elias JE, Gygi SP. Target-decoy search strategy for mass spectrometry-based proteomics. Methods Mol Biol Clifton NJ. 2010;604:55–71.

43. McAlister GC, Huttlin EL, Haas W, Ting L, Jedrychowski MP, Rogers JC, et al. Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. Anal Chem. 2012 Sep 4;84(17):7469–78.